20269 – Economics of European Integration

Take Home

Bucchi Filippo – 3186624
Fascione Luisa – 3187069
Manna Sem – 3087964
Pulvirenti Alessia – 3060894

Contents

Problem I

a.

Restricting the analysis to Italian firms in 2008 yields a cross-sectional dataset of 4,324. Of those, 3,277 (or 75.79%) concern observations for firms operating in in the textile industry (NACE rev.2 code 13) while the remaining 1,047 (24.21%) operate in the motor vehicles, trailers, and semi-trailers industry (NACE rev.2 code 29).

Looking at relevant variables of interest, we notice how the average real capital in 2008 of an Italian firm in the dataset is 1.12 million Euro, with a median of just 52,000 Euro and values ranging from 0 to 745 million. Moreover, given a standard deviation of over 15 million Euro, we note how capital varies vastly across firms, and its density has a large positive skew. Similarly, average real revenues amount to 14 million Euro, with a median of 2 million Euro and a standard deviation of 315,000 Euro. For half of the firms, we observe a real (deflated) value-added below 1.2 million Euro, with an overall mean value of around 5 million and hence a large positive skew. Similar considerations apply for the number of employees, which range from just 1 to 22,639. In 2008, the average Italian firm had almost 50 workers, while the median employed only 13 people. Hence, more than half of the firms analysed belong to the second category of the size-class variable, employing between 10 and 19 workers. This workforce costs firm an average of 1.6 million Euro per year, with a median of 393,000 and a maximum value of 905 million.

It is worth pointing out how firms in the dataset vary greatly across all relevant variables. The densities of the observed values show a large positive skewness, with very few observations displaying very high values. This can be noticed by looking at how the values of relevant covariates vary across percentiles. Starting from the 75th percentile, and especially after the 95th, values skyrocket as few observations display values further and further away from the median. This preliminary descriptive evidence, albeit relating to two industries only, is consistent with the established portrayal of the Italian economy as one composed of many small and medium-sized enterprises (SMEs) and few large – potentially multinational – companies, that indeed display much larger values of capital and of number of employees.

Restricting our analysis to the two industries separately, we point out how, on average, firms in the manufacture of textiles (or "industry 13", based on its NACE rev. 2 classification) are characterized by significantly smaller values than those operating in the manufacture of motors, trailers and semi-trailers ("industry 29") across all relevant dimensions reported in the data frame. Performing a t-test on the equality of the mean for each variable, the two industries display statistically significantly different averages for all relevant covariates.

For what concerns the number of workers, we observe that the means in the two industries are significantly different: 27.40 workers for firms in sector 13 versus 117.23 for firms in sector 29. As Figure I shows, this difference comes from the fact that the vast majority (around 65%) of firms in the textile sector employs less than 20 people, and over 40% has less than 9 employees. Moreover, it should be noted that, in sector 29, the number of workers of firms belonging to category 5 (i.e., those with 250+ employees) is considerably greater than that of firms of sector 13 belonging to the same size class, with mean values of around 1,600 versus 400. The percentile distribution of sector 29 points to the presence of an outlier (a firm with a labour force of more than 20,000 employees, 2.1 times greater than the largest right below it), but even after dropping that observation the difference is not significantly affected. The share of firms in sector 13 belonging to the largest category is approximately 1%, while it is above 5.3% in sector 29.

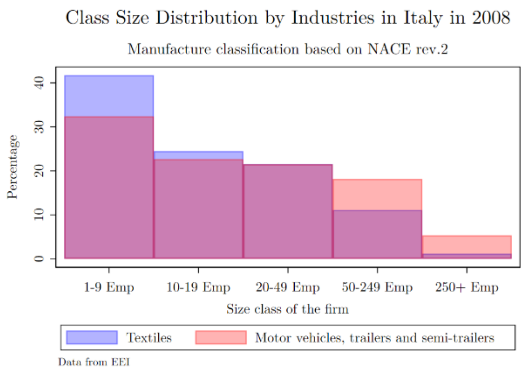


Figure I

Since the values of labour and other relevant variables are highly positively skewed and have a different range for sector 13 and for sector 29, we plot distribution of their logarithm to get a visual validation of their differences (Figure II). For instance, when looking at the number of workers, at lower values, the density of firms in sector 13 is above that of firms in sector 29 while, after a certain threshold, and hence for values corresponding to a higher number of workers, the density of firms in sector 29 is well above that of sector 13.

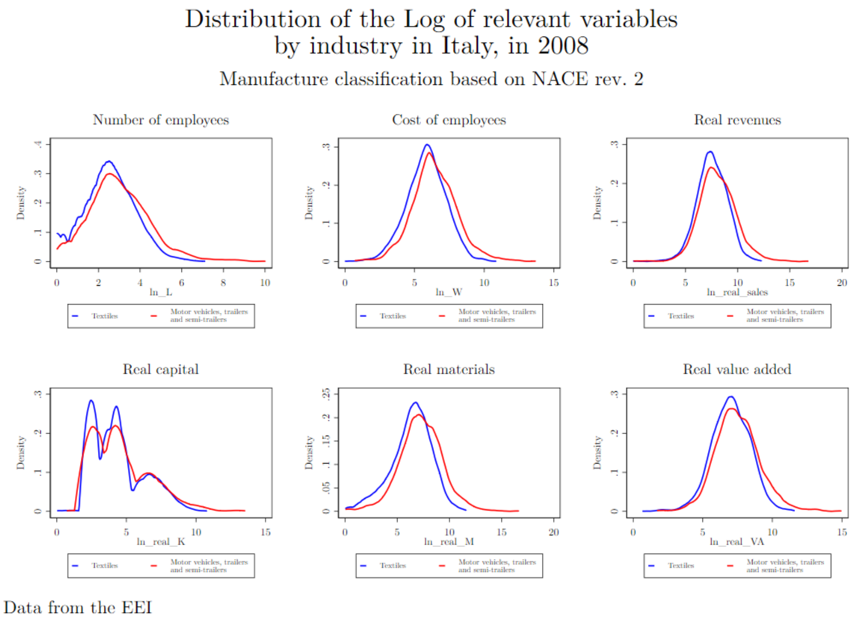


Figure II

For what concerns the accounting variables of the firm, we comment on deflated values rather than nominal ones, to control for price-inflation dynamics. The main intuition from Figure II appears to be that the density of number of employees, revenues, materials, and value added in Italy in 2008, for industry 29 (red line) is larger than that of industry 13 (blue line) for larger values of all variables. The only partial exception to this right-shift of the density in except real capital, whose visual interpretation is less clear as many observations are clustered around few values. Notably, values for real capital are clustered around values of (real) capital of 500 and 1000. Nevertheless, real capital as well as all the other variables plotted display a much longer right tail for industry 29, highlighting the presence of more large firms in motors than in textile manufacturing.

Table I presents summary statistics for the relevant variables in the two industries, highlighting their differences. The average of real sales in the textile sector amount to about 5.4 million Euro, versus 42 million in sector 29 – unsurprisingly, given the different nature of the businesses in the two sectors considered. Clearly, we would expect this discrepancy to be present also in the deflated values of capital and materials. Looking at material inputs, the average values of real material for sector 13 and sector 29 are respectively 2.5 and 33 million Euro.

As opposed to the significant discrepancy in revenues, the difference between the means of value added in the two sectors appears to be relatively smaller – possibly because material inputs, as measured by the value of raw materials, are significantly higher in the motor industry (sector 29) at all percentiles, thus offsetting higher sales values – with a mean of almost 12 million Euro for industry 29 versus around 2.8 million for sector 13.

For what concerns cost of workers, given the higher average number of employees in sector 29, we expect a coherently higher mean value of total wage costs per firm. Indeed, we observe mean values of around 900 thousand Euro in sector 13 versus 4 million for sector 29.

Table I

Summary Statistics for Italian Firms in 2008, by Sector

	Mean Sector 29	Mean Sector 13	Std Dev Sector 29	Std Dev Sector 13	Difference Means	Difference Std Dev
Number of employees	117.23	27.40	845.12	58.85	89.83	786
Cost of employees (k€)	4,118	919	32,200	2,253	3,199	29,947
Real revenues (k€)	42,093	5,164	637,960	11,836	36,929	626,124
Real capital (k€)	2,971	559	31,644	1,994	2,412	29,651
Real materials (k€)	32,775	2,492	559,054	6,657	30,284	552,397
Real value added (k€)	1,1981	2,797	108,844	6,071	9,185	102,773

b.

Restricting the dataset to Italy in 2017 yields a cross-section of 4,567 firms operating in the Italian textile and motors industries. The observations for sector 13 are 3,387 while for 29 are 1,173. We observe an overall slight increase in the number of firms which is similarly reflected in both sectors.

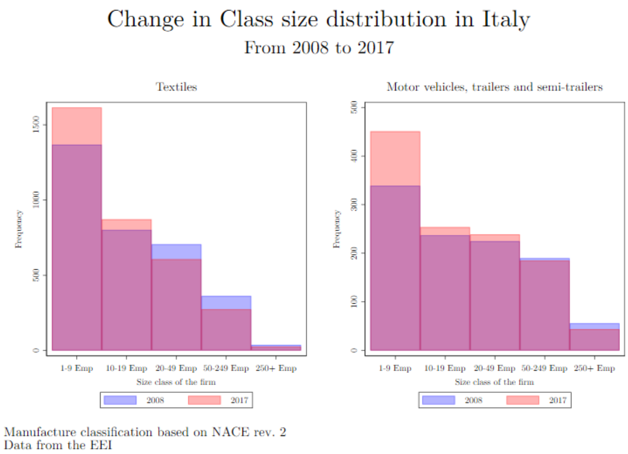


Figure III

The histogram in Figure III plots the frequency of observations in each class size by industry in 2009 and 2018. Notably, it can be seen how the number of small firms has increased for both sectors, while the frequency of larger firms has decreased substantially, especially in the textiles industry. We posit two explanations for this phenomenon. On the one hand, the decrease in the number of big firms could be due to the long-lasting/sticking effects of the financial crisis of 2008-2009 and to the sovereign debt crisis of the following years. In those periods, Italian firms were forced to reduce their workforce to cut on variable costs as the economy shrank. On the other hand, the increase in the number of firms in both sectors might be a consequence of the Italian Start-up Act, entered into force in February 2016. Indeed, as mentioned in the Annual Start-ups and SMEs report by the Italian Ministry of Economic Development, the number of start-ups in year 2017 has increased by 24,5%, as a result of the reforms apt at simplifying bureaucracy and cutting down setting up costs for new, innovative companies. This could potentially explain the small size of the potential new firms.

Table II

SIZE CLASS TRANSITION MATRIX (2008-2017) BY INDUSTRY, IN ITALY

Motor vehicles, trailers, and semi-trailers

Size Class	1	2	3	4	5	**Total**
1	12,926	648	29	6	2	13,611
	95.0%	4.8%	.2%	.0%	.0%	100.0%
2	629	6,529	349	6	0	7,513
	8.4%	86.9%	4.7%	.1%	.0%	100.0%
3	42	391	5,024	122	0	5,579
	.8%	7.0%	90.1%	2.2%	.0%	100.0%
4	6	12	160	2,544	9	2,731
	.2%	.4%	5.9%	93.2%	.3%	100.0%
5	3	0	0	14	219	236
	1.3%	.0%	.0%	5.9%	92.8%	100.0%
Total	13,606	7,580	5,562	2,692	230	29,670
	45.9%	25.6%	18.8%	9.1%	.8%	100.0%

Textiles

Size Class	1	2	3	4	5	**Total**
1	3,486	218	14	5	0	3,723
	93.6%	5.9%	.4%	.1%	.0%	100.0%
2	186	1,855	138	5	0	2,184
	8.5%	84.9%	6.3%	.2%	.0%	100.0%
3	18	104	1,780	89	1	1,992
	.9%	5.2%	89.4%	4.5%	.1%	100.0%
4	7	2	82	1,443	13	1,547
	.5%	.1%	5.3%	93.3%	.8%	100.0%
5	1	0	1	19	390	411
	.2%	.0%	.2%	4.6%	94.9%	100.0%
Total	3,698	2,179	2,015	1,561	404	9,857
	37.5%	22.1%	20.4%	15.8%	4.1%	100.0%

The transition matrix in Table II shows the cumulative changes in the number of firms in each size class from 2008 to 2017. Looking at percentage variations allows to identify the overall trend, while the frequencies with the number of firms is a cumulative result of the in-between transition matrices which does not reflect the final values in 2017.

The matrix confirms the decrease of firms belonging to size class 5 seen in Figure III: in both industries around 5% has moved to size class 4, both the industries presenting some extreme cases in which a very large firm has regressed to size class 1. Moreover, also for size class 3 (medium) and 4 (medium-large) it can be pointed out that the percentage of firms that have shifted to smaller dimensions is larger than those that have instead increased the number of workers. In particular, we highlight that in Motor vehicles, trailers, and semi-trailers industry 5.9% of firms has moved from size class 4 for to size class 3 and 7% from size class 3 to size class 2. In parallel, in Textile industry 5.3% firms has moved from size class 4 for to size class 3 and 5.2% from size class 3 to size class 2.

Such transitions seem to emphasize a trend directed towards a growth in the number of very small and small firms. In fact, in both industries almost all the firms that belonged to size class 1 in 2008 remain in size class 1 or moved to size class 2. Moreover, of the firms that were in size class 2 in 2008, only 4.7% (industry 29) and 6.3% (industry 13) moved upwards, to class 3.

Consistently, as reported in Table III, the average number of workers per firm decreases as well between 2008 and 2017. This holds both at aggregate level, moving from an average of 61 employees down to 55, and at the industry level, from an average of 27 in 2008 to 22 in 2017 for sector 13 and from 117 to 106 in sector 29. Indeed, we observe that the average value of total wage costs slightly decreases for sector 13, whereas it increases in sector 29 by 1 million Euro.

Table III
Summary Statistics for Italian Firms in 2008 and 2017, by Sector

Industry 29	Mean 2017	Mean 2008	Std Dev 2017	Std Dev 2008	Mean Diff	Std Dev Diff
Number of employees	106.08	117.23	1054.90	845.12	-11.14	210
Cost of employees	5,196	4,118	52,539	32,200	1,078	2,0339
Real revenues	51,886	42,093	815,496	637,960	9,793	177,536
Real capital	3,218	2,971	38,624	31,644	246	6,979
Real materials	36,239	32,775	604,852	559,054	3,463	45,798
Real value added	14,610	11,981	153,807	108,844	2,628	44,963

Industry 13	Mean 2017	Mean 2008	Std Dev 2017	Std Dev 2008	Mean Diff	Std Dev Diff
Number of employees	21.80	27.40	46.41	58.85	-5.61	-12
Cost of employees	856	919	2,136	2,253	-63	-116
Real revenues	4,240	5,164	12,012	11,836	-924	176
Real capital	398	560	1,928	1,993	-161	-65
Real materials	1,786	2,492	6,530	6,656	-705	-126
Real value added	2,407	2,797	5,784	6,071	-389	-286

The information conveyed by Table III yields an interesting picture for changes in Italian firms' performance from 2008 to 2017, in the aftermath of the financial crisis and European Debt crisis.

For what concerns sector 29, motors and vehicles, average (real) sales increase from 42 to 51 million Euro (or 23%), and so do average (real) material costs, but to a proportionally smaller extent from 32 to 36 million Euro, (or a 10% increase). Accordingly, profitability in the industry (proxied by real value added) increases, driven by higher sales volume produced at relatively lower average costs, going from 12 million to almost 15 million in 2017 (22% increase). If the average firm were actually representative of true industry dynamics, a potential argument may lead to pointing to higher economies of scale in sector 29 in 2017 with respect to 2008, since the increase in revenues is not offset by parallel increase in material costs. Moreover, largest firms have simpler access to international markets and they are able to exploit their network in order to obtain intermediate products at more competitive prices. In particular, considering the supply-side shock coming from imports from China, a further channel could probably account for the fact that, given an average lower wage level (especially of unskilled labour) and an average higher efficiency in production processes, intermediate inputs imported from China are found to be often much cheaper than domestically-produced ones, thus complementing the role of economies of scale and reducing material costs (Altomonte and Coali, 2020).

Looking at sector 13 instead, the trend appears to be different. Revenues in 2017 are lower than in 2008 by almost 20%, (real) material costs also decrease by more than 30% and (real) value added also decreases by 24% (from almost 2.8 million Euro to 2.4 million). This hints to what will be further discussed later on: the textile industry is among the most hit by import-competition from China in the manufacturing sector. Although China entered the WTO already in 2001, and the steep reduction of commercial tariffs made imports from China to western countries skyrocketed since then, the trend underlined in this paragraph might still hint to this phenomenon. For further comments on these dynamics, see Problem IV.

Problem II

a.

Table IV*Table IV* show the results of the production function coefficient estimation using different estimation procedures. We controlled for the fact that data come from different countries and years, and may thus be affected by time-invariant country-specific factors or time-variant common trends by including country fixed effects and year fixed effects. In this way, we both control for common macro-trends which vary over time and for time-invariant Country characteristics. Country Dummies generate entity-specific intercepts aimed at capturing time-constant heterogeneity across the subjects under analysis (in our case, countries).

Table IV

PRODUCTION FUNCTION COEFFICIENTS ESTIMATES

	(1)	(2)	(3)	(4)	(6)	(7)
Estimation Method	OLS	OLS	WRDG	WRDG	LP	LP
Industry	Nace 13	Nace 29	Nace 13	Nace 29	Nace 13	Nace 29
Outcome Variable	ln Real VA	ln Real VA	ln Real VA	ln Real VA	ln Real VA	ln Real VA
ln Labour	0.808***	0.911***	0.663***	0.683***	0.646***	0.646***
	(0.00230)	(0.00283)	(0.00182)	(0.00281)	(0.00825)	(0.0147)
ln Real Capital	0.164***	0.125***	0.0578***	0.0791***	0.0737***	0.0938***
	(0.00152)	(0.00198)	(0.00723)	(0.00806)	(0.0114)	(0.0155)
Observations	112,713	49,104	101,694	44,714	112,176	48,942
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Number of groups			9,888	4,038		

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

b.

The bias on the beta coefficient for the logarithm of labour arises from the so-called "transmission bias", first identified by Marschak and Andrews (1944). This key challenge in identification stems from noticing that firms' optimal input decisions may follow from their productivity levels (unobserved in the data), rather than the other way around (Gandhi et al., 2020).

Moreover, this would also induce a bias in the Total Factor Productivity (TFP) estimate. Recalling how the TFP is derived from the Solow residual of the firm-level OLS predicted output minus the observed one, within the context of a Cobb-Douglas production function, we should reckon how inputs decisions and shocks can be jointly determined. Since we measure productivity from year-end balance sheet data, this does not allow to capture infra-year potential adjustments to shocks, leading to a correlation between our TFP estimate (the residual) and the inputs (the independent variables). This *simultaneity* issue leads to biased OLS estimates which, if firm-specific, could be accounted for using firm fixed effects.

The rationale behind the bias in both the labour and TFP estimates lies in the fact that, following a productivity shock, the firm will likely attempt to exploit such productivity shock by hiring a higher number of workers. From an econometric standpoint, this makes the error term and the independent variable (labour), correlated, violating the exogeneity assumption of OLS. Typically, the bias of the labour coefficient will be upward, thus also biasing upward the predicted output on the one hand and underestimating the TFP on the other. Additionally, since the estimated TFP will be positively correlated with one of the variables, it cannot be considered as a proper error term but rather almost as an omitted variable.

The bias on the labour coefficient can result in three different realizations. When labour is the only input decision responding to the shock (i.e., more workers are hired following a productivity shock), and labour is uncorrelated to capital, then the coefficient on labour will tend to be biased upward, but the estimate for capital will remain unbiased. If instead capital and labour are positively correlated, the estimate for capital may, in addition, present a negative bias. Lastly, if capital and labour are positively correlated but the labour's correlation with the productivity shock is higher than capital's correlation, the model will tend to overestimate labour and underestimate capital (Levinshon and Petrin, 2003).

A first solution is using firm-specific fixed effects in the estimation of the OLS, which solves the issue of simultaneity by controlling for plant-specific but time invariant characteristics (Mundlak, 1961; Hoch, 1962). This allows to study the within-firms variation in productivity, but will still lead, at best, to weakly identified OLS coefficients. Indeed, the methodology assumes that unobserved productivity does not vary over time. In the process of estimating the TFP, it makes limited sense to exclude events related to TFP shocks, and the methodology is inconsistent with the need to capture productivity patterns of business cycles. Additionally, the fixed effects estimator assumes that the unobserved component of the error term that is causing the bias is fixed over time, while the input endogeneity can likely exist with respect to a time varying unobservable. Additionally, weak identification also comes from recognizing that the key variation exploited by a FE model after clearing for time and country fixed effects is within-subjects variation, here within-firms, which in micro-data tends to be much lower than the cross-sectional one.

The bias on the OLS coefficient for labour is addressed by more complex models for estimating the production function and the TFP. We will provide a quick overview of the main ones.

First of all, the upward bias on all coefficients might also be traced back to the perhaps unavoidable "*selection bias*" resulting from only accounting for surviving firms – i.e., those with a positive output (that remain active) on the whole timeframe considered in the sample. A first attempt to address both biases (simultaneity and selection) comes from Olley and Pakes (1996), who posit that incumbent firms decide at the beginning of each period whether to continue participating in the market or to exit. Indeed, the dynamic model that they propose allows for firms' selection by including firm-specific change dimensions and industry turnover. The key feature of the methodology is using investment decisions dynamics to account for unobserved productivity shocks, $\omega_{i,t}$ (of which investments are assumed to be an increasing function). This adds a step in the estimation, whereby the probability of exit is estimated non-parametrically in a two-stage model (Olley and Pakes, 1996).

An expansion and improvement of this methodology is Levinshon and Petrin's regression (2003), which exploits again a semi-parametric technique that takes TFP as a function of capital and materials. The methodology expands upon Olley and Pakes' framework, using intermediate inputs (instead of investment decisions) as instrumental variable and defining productivity as a function of them and of capital.

Specifically, the LP methodology relies on estimating a demand function for inputs that is a function of productivity and capital and, assuming its monotonicity and invertibility, expressing productivity as a function of intermediates and capital. This expression is then used in an auxiliary function, used to estimate a first stage of the regression with output as a function of labour and of the auxiliary, to estimate the labour coefficient. The coefficients on capital and materials are then estimated in a second stage (Levinshon and Petrin, 2003).

Ackerberg, Caves and Frazer (2015) find evidence of collinearity issues in the first stage of LP estimation, showing that the coefficient of labour is still biased by dependence on the level of inputs (materials). The ACF corrected model takes into account labour market rigidities assuming materials as a function of labour as well (in addition to capital and productivity) (Ackerberg et al., 2015).

A last TFP estimation procedure, introduced by Wooldridge and regarded as a final step forward from the aforementioned methodologies, implements the proposed two-steps procedure through a Generalized Method of Moments (GMM) model. The methodology also allows to overcome some multicollinearity and identification issues that are still present in the ACF-corrected LP framework. Nevertheless, the methodological framework of the LP and WRDG procedures is fairly similar, which explains why the values of the bias on the OLS coefficient compared to these two models are close, as Table V shows. Finally, the WRDG approach allows to efficiently estimate robust standard errors.

We indeed assess the bias in the OLS coefficients of our production function estimation by taking into account both its discrepancy from both LP and Wooldridge estimates.

Table V

TFP Coefficients by Industry and Estimation Procedure

		Nace-13	***Nace-29***
Lev-Pet	ln(labor)	0.646***	0.646***
	ln(capital)	0.0737***	0.0938***
WRDG	ln(labor)	0.663***	0.683***
	ln(capital)	0.0578***	0.0791***
OLS	ln(labor)	0.808***	0.911***
	ln(capital)	0.164***	0.125***
	Bias in labor coefficient - with respect to LP	0.162	0.265
	Bias in labor coefficient - with respect to WRDG	0.145	0.228
	N. of observations (OLS)	112,713	49,104

Table V confirms the intuition that the OLS coefficient for labour has an upward bias when compared to coefficients estimated using both LP and the Wooldridge procedures. Notably, the amount of the bias with respect to the two estimations is quite close.

Problem III

a.

In the previous problem we have estimated the so-called *value-added production function*, using the variable value added, already provided in the data set, to produce an estimate of the coefficients of labour and capital, and of TFP. However, we could also have estimated a *turn-over production function*, i.e., one which uses turnover/sales as outcome variables. This is reasonable, considering that our theorical production function is a Cobb-Douglas:

$$Y_{it} = A_{it} {L_{it}}^{\alpha} K_{it}^{\beta} M_{it}^{\gamma}$$

Where;

- Y_{it} is the measure of total sales (rather than Value Added),

- *Intermediate Materials* is included on the right-hand side as a factor of production as M_{it},
- A_{it}, the inputs' multiplier, is an unobservable component regarded as a proxy for technological progress, defined in the literature as $A_{it} = C\varepsilon_{it}$, i.e., determined both by a deterministic (constant) component and a random term
- Parameters α, β, γ relate to returns to scale.

Therefore, estimating a turn-over production function, log-linearising the Cobb-Douglas, would entail estimating:

$$y_{it} = a_{it} + \alpha l_{it} + \beta k_{it} + \gamma M_{it} \tag{1}$$

Where the multiplier is now $a_{it} = logA_{it} = c + \varepsilon_{it}$ is composed of the model's intercept c, representing the mean efficiency level across firms over time, and the error term ε_{it}, time- and firm-specific deviation from that mean. The latter term can generally be further decomposed into two components, an observable (ω_{it}) and an i.i.d. unobservable one (u_{it}), which leaves the model equation as:

$$y_{it} = c + \alpha l_{it} + \beta k_{it} + \gamma M_{it} + \omega_{it} + u_{it} \tag{2}$$

The resulting equation (1) is now a function with three time-varying parameters (α, β, γ), rather than two, to be estimated, and ω_{it} is firm-level productivity (TFP). The log-linear function allows to bring the TFP on the left-hand side, and estimate productivity levels as $\Omega_{it} = \exp(\omega_{it})$.

The main limitation of the assumptions behind the Cobb-Douglas production function is the arbitrary level of substitution of factors of production. For example, a lower number of workers could be compensated by a higher amount of capital, obtaining the same value of sales. This assumption works reasonably for the substitution between labour and capital, but it is hard to conceive as plausible when thinking about substitution of labour with material. Consequently, the unrealistic perfect substitutability of factors posited by the Cobb-Douglas model might hinder the validity of the identification of the coefficients of equation (2) above.

Given the linearity of the log-linearised Cobb-Douglas, the Materials term can be brought to the left-hand side of equation (2) and subtracted to sales, obtaining value added as the preferred outcome variable. This procedure, on top of preventing theoretical flaws in the specification, allows to work with a more streamlined model, with only two coefficients to be estimated.

A further point to be touched upon when discussing the use of sales in computing factors' productivity – which is supposed to measure the efficiency of an input – is that it would also consider (on the left-hand side) value actually created outside the firm by third parties, i.e., the supplying firms from which the intermediate materials are purchased. Consequently, the Total Factors Productivity estimated through revenues, instead of value added, may likely distort the true value of a firm's productivity. The value-added approach is generally considered as having sizeable advantages stemming from ignoring the difficulties of dealing with inter-industry and intra-industry flows of (intermediate) good and services, making measures more reliable as well as comparable across sectors and industries (Cobbold, 2003).

Needless to say, estimating the turn-over production function omitting Materials as one of the independent variables, would generate biased estimates of our coefficients due to Omitted Variable Bias (OVB). Using revenues rather than value added, will likely affect the estimates of all the three methods. As mentioned, the OLS specification will suffer an omitted variable bias, which is very likely to be an upward bias both for labour and capital. If we rely on Levinshon-Petrin, instead, it is enough to include the option revenues, which tells Stata not to use the simplified version of the polynomial expansion to estimate the demand for capital and material, $\phi_t(k_t, m_t)$ (Problem II).

In practice, when performing the estimation using sales rather than value added, the OLS estimates appear to underestimate the coefficient of labour (down to almost half the value of estimate found using LP and Wooldridge) for both sectors, and a similar impact is found for the coefficient of capital, as including materials on the right-hand side captures some of the explanatory power of the other variables.

The degree of openness of the overall economy in the setting considered is also found to be a determinant of the discrepancy between productivity measures using real revenues and those using added values. Indeed, it is in general observed that in a closed economy the absolute value of their difference decreases as the level of aggregation considered increases. Indeed, at the total economy level, the values of the productivity estimate based on value-added based and on gross output asymptotically coincide. A discrepancy still arises, however, in the presence of imports (open economy), whereby the two measures produce different results even at total economy level (Schreyer 2001, p. 42). In our framework of analysis therefore, the distinction plays a key role.

Finally, according to OECD definitions, the value-added measure of productivity growth shall not be considered as a measure of technological change in an industry or a measure of overall improvements in efficiency. Rather, it shall be looked at as an industry's capacity to translate technological change into income and a contribution to final demand (OECD 2001, p. 25). That is, it reflects changes in an industry's contribution to aggregate income.

Problem IV

a.

Table VI

Percentile Distribution of the OLS-computed TFP for industries 13 and 29

TFP (OLS) – Textiles				
	Percentiles	Smallest		
1%	41.32617	36.48357		
5%	77.9822	36.48357		
10%	113.8431	36.48357		
25%	251.321	36.48357	Obs	119,170
50%	653.9939	Largest	Mean	1,845.228
75%	1,643.855	143,062	Std. dev	4,459.332
90%	4,170.488	159,887	Variance	1.99e+07
95%	7,315.641	161,563	Skewness	9.914242
99%	19,207.56	181,207	Kurtosis	179.5107

TFP (OLS) – Motor vehicles, trailers, and semi-trailer				
	Percentiles	Smallest		
1%	52.69734	39.80228		
5%	102.0692	39.80228		
10%	174.7164	39.80228		
25%	422.8643	39.80228	Obs	51,250
50%	1178.473	Largest	Mean	13255.33
75%	4,373.642	5,185,893	Std. dev	118785.8
90%	15,891.19	5,362,597	Variance	1.41e+10
95%	33,328.35	5,547,195	Skewness	29.87261
99%	219,022	5,569,692	Kurtosis	1083.374

Table VI reports summary statistics for the OLS-Computed TFPs (before cleaning for outliers) for France, Italy, and Spain, by industry. Concerning sector 13, we note that, given the median value of 653.9 the firm corresponding to 95^{th} percentile is 1,020% more productive while the 99^{th} one is 2,084% more productive. The increment is significant but plausible, what appear extreme are the largest values presented in the distribution summary. In fact, a value of 19,207 in the 99^{th} percentile is followed by values such as 159,888; 161,563; 181,207 in the very last part of the tail. This is a clear signal of the presence of outliers in the distribution of the OLS-computed TFP in the textile industry.

For what concerns the TFP distribution of sector 29, also in this case we observe that the largest values of TFP is too large with respect to the 99^{th} percentile. Consequently, we can assume the presence of outliers and we expect obviously a reduction in standard deviation which will lead to a decrease in mean of the OLS-computed TFP in motor vehicles industry. If we restrict the range to 1st-99th percentiles we would expect a reduction in standard deviation and also a decrease of TFP mean. In fact, we will observe that the new largest observation (219,022) will be consistent (almost 150% increase) with the post-cleaning 99^{th} percentile's value (88,021.55).

After cleaning, we note that now in both the distributions the largest values seem to follow a consistent path if compared to previous percentiles' values. As we expected the standard deviation decreases and also the mean does the same, confirming the presence of outliers in the original TFP distributions.

Figure IV shows a comparison of the outputs of the estimation methods, the densities of the estimates computed through LevPet and Wooldridge are almost overlapping, with the average being greater in sector 13 than in sector 29.

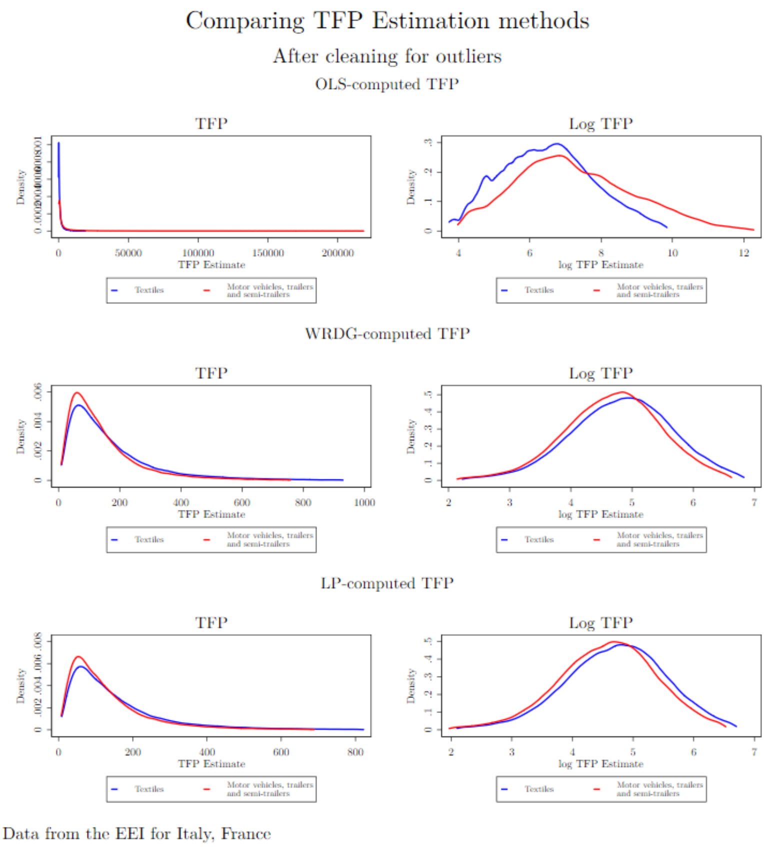


Figure IV

After the cleaning for outliers, the TFP distribution obtained through OLS method presents a more Pareto-shaped distribution, if compared with the uncleaned residuals distribution. Nevertheless, the plot appears heavily left concentrated due to the long right tail of sector 29's distribution. For this reason, we provide comments focusing on the corresponding Log TFP plot.

Relying on simple OLS estimation, the distributions for both sectors appear fairly unregular and Log distributions are distorted with respect to what is presented in LP or WRDG procedure. In this latter case, productivity levels in the two sectors are inverted with a higher concentration of high TFP firms for sector 29 and consequently a higher concentration of low TFP firms for sector 13. The difference between the graphs produced with OLS and Lev-Pet or Wooldridge can be traced back to the discussion on the underestimation of the residuals with the OLS method, provided in the last paragraph of problem II.

b.

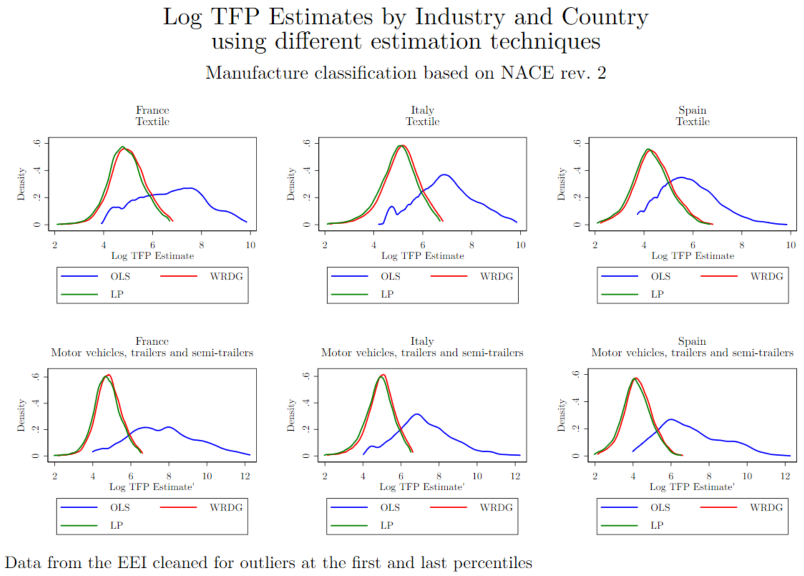


Figure V

Figure V shows the densities of the log TFP distributions of Italy, France and Spain, by industry, for methodological comparison. We chose to plot the logarithms for an easier comparison of the different methodologies throughout the sample. As noted in previous points, the kdensities show that, for each country and industry, the TFP distributions obtained with LevPet and Wrdg methods almost overlap, while the OLS-estimated one has a much less regular shape and is shifted to the right, towards higher TFP values. The OLS seems to overestimate the contribution of Total Factor Productivity to output growth.

Figure VI plots the densities of the distribution of the Log TFP distributions of Italy, France and Spain, emphasizing cross-country differences as well as within-country differences across sectors, for each estimation method. The solid lines show values the 5th and 95th percentiles.

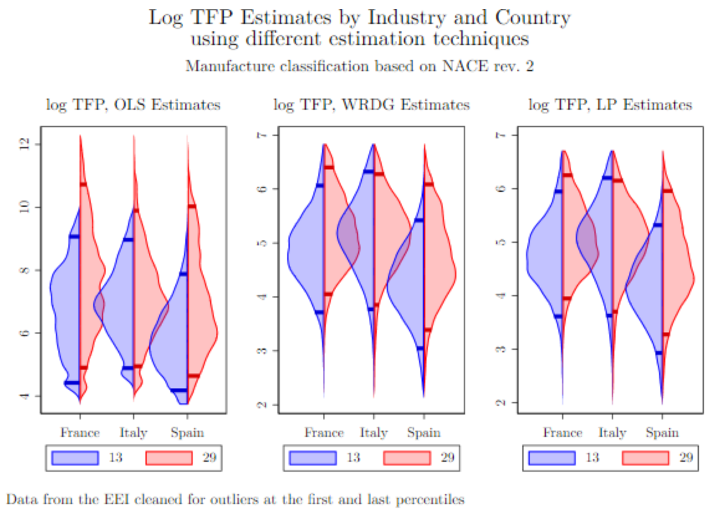


Figure VI

The violin plots are a different representation of the densities in Figure V above, which plot the two sectors together and allow cross-country comparisons. In particular, we observe that, according to LP and WRDG estimates, France and Italy's distributions for both sectors appear more aligned than with respect to Spain's, whose density appears to be shifted towards lower values of TFP estimates throughout the sample. Notably, Italy's Textiles industry shows a smaller TFP wedge with the Motors one, when comparing the values at the 95th percentile, then in France and Spain.

c.

Focussing our attention on industry 29, for Italy and France. To begin with, it is worth noting that in Italy the number firms active in sector 29 in 2008 increased from 2001 (1,047 vs 975), while it decreased in France (789 vs 843). Given that the changes are not extremely significant, and many confounders could be at play, we cannot attribute to the China Shock a considerable impact on the industry's dynamics without performing a micro-level analysis of the industry's turnover (entry and exit, potential mergers and acquisitions…).

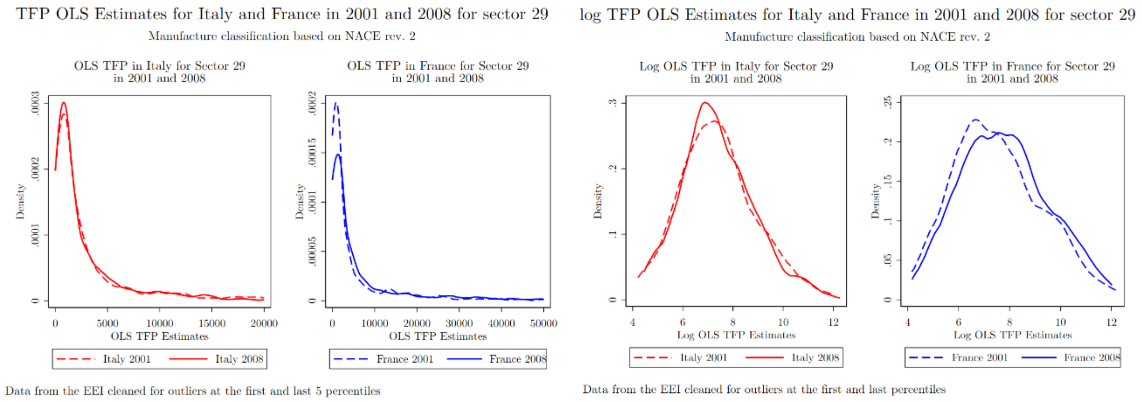


Figure VII

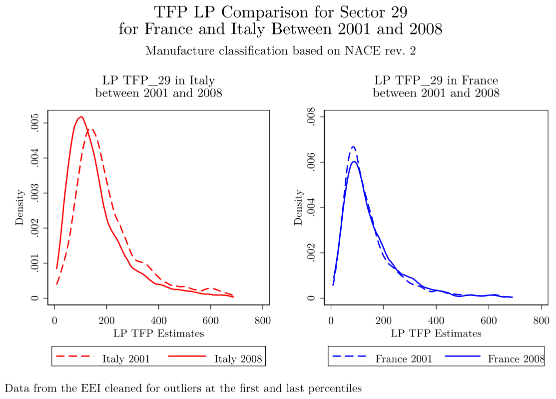
For what concerns the changes in the OLS-computed TFP distributions, Figure VII and Table VII seem to show a consistent picture. France presents a greater average TFP in 2008 than in 2001 (around 14% increase) coupled with a higher median value and a negative change in skewness, pointing to an overall higher density of larger TFP values. Italy, on the other hand, presents a lower average TFP in 2008 (less than 5% decrease) and positive skewness change, indicating a higher number of firms with lower TFP values compared to 2001.

Table VII

Changes in the OLS-Computed TFP in France and Italy for Sector 29 (2008-2001)

	Mean 2001	**Mean 2008**	**Mean Diff**	**Skw 2001**	**Skw 2008**	**Skw Diff**
Italy	5766.306	5502.843	-263.463	6.081138	7.403116	1.321978
France	8813.364	10068.99	1255.626	5.454621	3.851017	-1.6036

Because we have argued that TFP estimations under OLS are overall biased, we inquire the Levinshon-Petrin and Wooldridge estimates for comparison of the results.



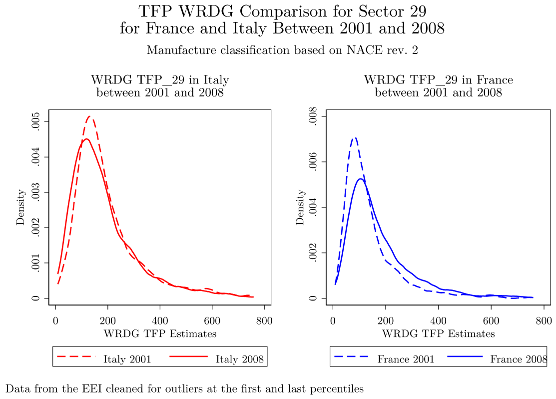


Figure VIII

Looking at the shift on the graph on the left in Figure VIII, we note a decrease in the average of Italian productivity of sector 29 estimated through Levisohn-Petrin. From 2008, average TFP drops by around 20% (a greater percent change than under OLS), and the same trend is observed, although to a lesser extent (around 5% negative change), in the TFP obtained through the Wooldridge procedure (on the right of figure Figure VIII). Table VIII reports summary statistics for both years, along with the change from 2001 to 2008.

Looking at France, the graph of the density of the TFP estimated through LP seems to remain fairly unchanged, and only from its percentile distribution we observe a slight and non-significant increase (around 3%) in the mean from 2001 to 2008. The pattern of change is less clear from the graph on the right (Wooldridge), but by looking at Table VIII we can see how the mean increases from 2001 to 2008 by almost 25%, much more than under LP.

Being this an exercise of simple means comparison (without adding controls), it is hard to give a proper interpretation of the observed changes in TFP. One hypothesis we can posit is the one stated by Altomonte and Coali (2020), whereby countries mostly affected by the China shock, in the medium term, have a lower TFP growth (despite having a higher TFP overall). This can be due by lags in the adjustment of the labour force and of the investments necessary to counteract the effects of the shock. If this hypothesis applies to our case, then Italy would be the country relatively more affected by the shock in industry 29. Another hypothesis could be advanced: following the entrance of China in the WTO and the surge of imports from this country, in the medium-run France and Italy have responded differently to the shock, with Italy being slower in the reallocation process, where bigger firms increase their productivity and smaller firms exits the market. Nonetheless, the literature on the topic has established that higher TFP is associated with higher intensity of the weighted import shock (Bloom et al., 2016).

Our conclusions could be corroborated only through an empirical test, with country-specific controls, to quantify the magnitude of the shock in both the countries in the analysed industry, especially in light of the fact that sector 29 does not belong to the manufacturing industry that the literature presents as the most vulnerable to China's imports raise in Europe (Colantone and Stanig, 2018). Carrying out this test would require having more granular data on the nature of Chinese manufacturing imports in each European country.

d.

Table VIII

CHANGES IN THE TFP DISTRIBUTIONS IN ITALY AND FRANCE (2001-2008)

	Mean 2001	Mean 2008	Diff Mean	Skw 2001	Skw 2008	Diff Skw
LP						
Italy	202.9936	162.0601	**-40.9336**	1.413012	1.61425	**0.201238**
France	147.3266	151.6981	**4.371497**	2.011662	1.831726	**-0.17994**
WRDG						
Italy	197.7115	187.138	**-10.5734**	1.627791	1.520035	**-0.10776**
France	140.5817	174.8409	**34.25924**	2.208177	1.723606	**-0.48457**

We recall that positive skewness indicates that the distribution has a long right tail, implying that higher densities correspond to relatively lower values on the x-axis (here TFP). Vice versa, negative skewness indicates a long-left tail, indicating higher density on larger values of the x-axis. Indeed, a negative change in skewness is reflected in higher density moving to the right (corresponding to higher TFP values). This would suggest a reallocation of production factors in favour of the “happy few”, i.e., the (small) share of the labour force who gain from the supply side shock. An increase in skewness values indicate a less dispersed distribution, with relatively higher number of small and unproductive firms, with significant room for redistribution and more efficient resources allocation.

Given the previous discussion, Table VIII indicates that sample skewness (obtained through the command `sum, d` in Stata) in 2008 decreases for France under both estimation methods, by around 10% under LP and 20% under WRDG. Visually, this would point to a distribution less skewed to the right, as Figure VIII indeed confirms (especially in the graph on the right).

The evidence for Italy from the two estimation procedures instead is not clear-cut, in fact LP would indicate a positive change of around 14%, while WRDG shows a decrease of 12%. The difference between the methodologies is substantial, and we decide to comment on what seem the most conservative measures provided obtained through the WRDG method. This was not as visually clear by looking at the shapes of the distributions in Figure VIII.

A trade shock leading to reallocation of market shares toward more productive firms should yield decreasing skewness parameter. If we rely on WRDG procedure (for the reasons outlined in the discussion of section IIb), Table VIII seems to show evidence in this direction both for Italy and for France. Such argument suggests that both countries were significantly affected by the China shock in a way that changed within-industry dynamics in favour of more efficient firms, in France more (20% decrease) than in Italy (12% decrease), for the motor vehicle industry. This is consistent with the hypotheses put forward in point c. regarding changes in average values of the two TFPs.

e.

In order to rigorously assess the magnitude and nature of the shifts in TFP in industry 29 across years 2001 and 2008, we rely on one estimation method of the skewness of the densities, which will return more accurate densities than the ones in Table VIII. For this last part of the analysis, we rely exclusively on the Wooldridge methodology for the estimation of TFP.

Looking at the right-hand side of Figure VIII, the best candidate for the distribution is a Pareto one. The cumulative distribution function of a Pareto-distributed variable is $F(X) = 1 - \left(\frac{X}{X_m}\right)^{-k}$, where X_m represents the lowest level of TFP attained in the observed sample (scale parameter), and $-k$ is the shape parameter of the distribution. This parameter represents the skewness of the population distribution, and, together with the scale parameter, can be estimated to provide consistent parametrical measures of the distributions under investigation. We proceed to estimate the scale parameter, $-k$. In particular, following the work of Norman, Kotz and Balakrishnan (1994), we perform an OLS estimate of k. This is done by regressing $ln(1 - F(X))$ on $ln(X)$ with a constant, and obtaining the estimate of the shape parameter by looking at the estimation of the slope of the regression line, $\hat{\beta}_1$. The produced estimates, performed in lines 1268-1290 of our do-file can be summarised in the following table:

Table IX

OLS ESTIMATION RESULTS OF THE SKEWNESS – (NORMAN ET AL. 1994)

	Skw 2001	C.I. 2001	Skw 2008	C.I. 2008	Diff. skw
Italy	1.329	[1.353573; 1.303745]	1.192	[1.214231;1.168929]	-0.137*
France	1.363	[1.390285; 1.336137]	1.308	[1.341831; 1.274778]	-0.055*

The R-squared of these regressions is, as expected very high. These estimates are then used, through a Kolmogorov-Smirnov test, to rigorously prove that our distributions are Pareto-distributed. The test yields a maximum significance corroborating the assumption made by looking at the shape of the graphs. Table IX suggests that the change in skewness is homogeneous – it goes in the same direction in both countries in the Manufacture of motor vehicles industry – although the delta differ sizeably in magnitude. The decrease is significant for both countries, although Italy experiences a higher decrease in skewness of its TFP distribution in industry 29, for year 2008. This consistent estimate of the decrease in skewness is not very noticeable in the if we look at the right panel in Figure VIII, which uses the WRDG estimation methodology, proving that the estimate provided "graphically" by looking at the sample data is not perfectly informative. We recall that the estimation method used allows to have a consistent (better) estimation of the skewness in the distribution of the population of the firms in industry 29 in Italy and France. The results found are in line with the hypothesis that both the countries have experienced a trade shock in this industry. In particular, it seems that Italy has experienced a stronger negative change in the TFP distribution (with a more left-skewed distribution), and this would be in line with one of the hypotheses posited in section c., which sees Italy as potentially more affected by the China shock. Indeed, Altomonte and Coali (2020) find that TFP distribution in regions more exposed to the China shock changed in the direction of a more left-skewed one. Again, as reported in section c., a regression with mean skew as outcome variable with proper country-specific controls and country fixed effects would shed light on the actual effect of the China shock, which in this preliminary analysis of comparison of mean outcomes could be contaminated by omitted variable bias.

Finally, the significance of the change in skewness is further verified through a Kolmogorov-Smirov test where, using the skewness estimated with the above procedure, we are able to confirm that in both countries, the distribution of the TFP in 2001 is statistically different from that in 2008.

Problem V

a. (See do-file)

b. (See do-file)

c.

The two maps, presented in Figure IX below, appear very similar. To a good extent, this is expected since the China Shock in a given NUTS-2 region,

$$ChinaShock_{crt} = \sum_j \frac{L_{rj}(pre\text{-}sample)}{L_r(pre\text{-}sample)} \cdot \frac{\Delta IMPChina_{cjt}}{L_{cj}(pre\text{-}sample)}$$

is computed by normalising the change in real imports from China in each industry by the total number of workers in the same industry in the whole country at the beginning of the sample period. Then, this measure of the change in imports per worker across industry is weighted by the relative importance of that industry in a given NUTS-2 region. Finally, the weighted measure obtained is summed across all industries to obtain a weighted average of the level of exposure to imports from China in that region, across the whole economy (all the industries). As a consequence of the way it is constructed, the China Shock measure will be stronger in those regions where a larger share of workers (in pre-sample year, hence exogenous to the shock) was employed in the industries most vulnerable to import competition from China (Colantone, Stanig, 2018 APSR). The reason behind using employment rates in pre-sample year (i.e., before the shock occurs) can be traced back to the endogeneity of the composition of the workforce to the shock itself. In other words, weighting the import shock from China by the number of workers in an industry which has already been hit by a shock, will, very likely, provide a flawed measure of the China shock, as workers might have already moved to another sector, or been fired, as a consequence of the import shock itself. The import shock from China is a phenomenon related to the surge in trade between industrialised economies and China in the manufacturing sector, with China becoming the world leader in the production of manufacturing goods, especially after its official entrance in the WTO in 2001. Consequently, the noticeable overlapping between the two maps in Figure IX was expected. The imperfect overlapping of the maps in some regions, for example in Piedmont, Lorraine and Champagne-Ardenne, where the highest share of employees in the manufacturing sector does not match the highest levels of China Shock, might be due to within-sector (manufacturing) heterogeneity in imports from China. Indeed, as Colantone and Stanig (2018, AJPS) note, the industries which experienced the greatest surge in import competition from China in Europe are, among others, manufacture of leather and leather products (22.96% of total imports in 2006), textiles (17.5%) and electrical equipment (13.21%) (Table 3A of SI). Therefore, given the same share of employment in the whole manufacturing sector, a region can be more or less affected than others depending on the composition of its production in this sector.

Comparison between 5-year average China shock years 1989-2006 and
Employment in manufacturing in pre-sample year

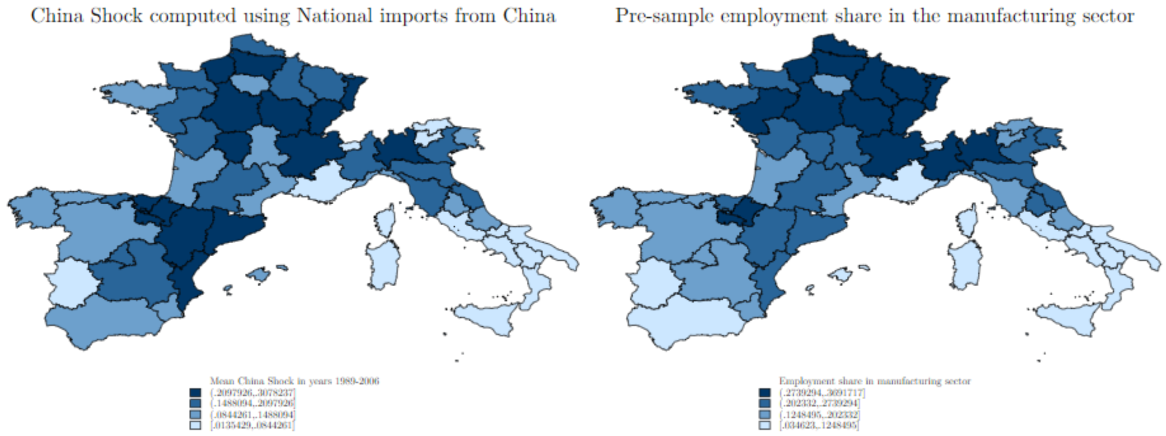


Figure IX

Problem VI

a.

Column (1) of Table X reports the results of a linear regression model absorbing Country and industry Fixed Effects. Including relevant controls, this OLS regression of the average post crisis (2014-2017) industry-region TFP on the average regional-level China Shock (1995-2006) shows a coefficient of 1.288, significant at the 99% confidence level. Multiplying this coefficient for the standard deviation of the China Shock variable we get the perhaps more interpretable figure of the impact of one standard deviation increase in the China Shock on the average TFP: 0.1012, or 4.52% of the mean of the average TFP in the sample.

Table X
Industry-Region Estimates:
Impact of the Average China Shock (1995-2006) on Post-Crisis Outcomes

	Average TFP (2014-2017))		**Average Wages (2014-2017)**			
	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Method	OLS	2SLS	OLS	2SLS	OLS	2SLS
Dependent Variable	**Avg. TFP**	**Avg. TFP**	**Avg. Wages**	**Avg. Wages**	**Avg. Wages**	**Avg. Wages**
China Shock	1.288***	1.267***	34.22***	32.49***	7.455	5.914
	(0.162)	(0.157)	(5.387)	(5.447)	(4.495)	(4.178)
Log Population (2014)	-0.00766	-0.00744	-0.817	-0.798	-0.691*	-0.677*
	(0.0131)	(0.0130)	(0.519)	(0.513)	(0.407)	(0.400)
Share with Tertiary	0.144	0.149	19.11**	19.56**	14.34***	14.55***
Educ (2014)	(0.237)	(0.238)	(7.837)	(7.962)	(4.728)	(4.792)
GDP Control	1.031***	1.029***	21.64**	21.48**	1.088	0.612
	(0.212)	(0.212)	(8.154)	(8.240)	(6.271)	(6.342)
Mean TFP					20.43***	20.81***
					(2.021)	(2.133)
Mean Dependent Var	2.230	2.230	37.71	37.71	37.71	37.71
Observations	640	640	643	643	640	640
R-squared	0.904	0.372	0.730	0.134	0.772	0.256
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
First-Stage Results						
IV China Shock	-	0.0578***	-	0.0579***	-	0.0554***
	-	(0.00289)	-	(0.00290)	-	(0.00308)
F-statistic instruments		400.7		398.6		323.7

Robust standard errors in parentheses, clustered at the Nuts2 level
*** p<0.01, ** p<0.05, * p<0.1

Trade theory offers many explanations on why import competition may increase the average TFP in a Country. Melitz's model of trade shows that the intra-industry effects of import competition can lead to higher average TFP through a reduction in the number of domestic firms operating in the market (Melitz, 2003). The positive estimated coefficient could thus result from *reallocation effects*, as import competition drives prices down, forcing less productive firms out of the market. The resulting market equilibrium will thus be characterized by the surviving firms – those with productivity levels high enough to stay in the market. The Pareto-shaped TFP distribution will still be characterized by many firms operating at the domestic level, clustered at relatively low levels of productivity, with now few ultra-productive multinational firms at the higher end of the distribution, with higher post-shock average TFP where firms had been more affected (i.e., in the regions more exposed to import-competition from China due to historical specialization). Similar conclusions can be reached if we consider the 2008 model by Melitz and Ottaviano and think of the China Shock as an enlargement in the market size, leading to an increase in the competitive forces of the market and thus survival of the fittest, most productive, firms in ever expanding markets. (Meliz and Ottaviano, 2008).

The OLS estimated coefficient in Table X is thus a positive and quite sizable one, yet its causal interpretation is not straightforward. One concern, as in Autor et al. (2013) is that the measure of import penetration from China may be correlated with import-demand shocks in the three Countries. The OLS estimate may then be biased if National import shocks, underpinning the China Shock index, are driven by (endogenous) domestic demand shocks. Such domestic demands factors, perhaps unobserved, could in turn drive both TFP growth and demand for Chinese imports, leading to an upwards bias in the OLS estimate. Then, if demand-side factors drive imports from China, the OLS estimate will be biased by endogeneity, and resorting to an exogenous instrument allows to identify and isolate the exogenous part of the China shock, to estimate its impact on domestic outcomes.

Following Autor et. al (2013) identification strategy and its application to the European context in Colantone and Stanig (2018), we use import penetration from China to the US as an instrument to predict imports in France, Italy, and Spain. The instrument is then built analogously to the original China Shock index but using Chinese imports in the US rather than in the three European Countries, an maintaining pre-sample regional employment shares to net out concerns related to simultaneity bias.

The first requirement for the instrument to be valid, is that it must have a relevant first stage, that is, it must be strongly correlated with the endogenous regressor to be instrumented. This can be checked by looking at the F-Statistic of the First stage.

The instrument must then be exogenous. This assumption is untestable and twofold. First, it requires that the instrument is as good as randomly assigned, and so uncorrelated with potential outcomes conditional on the observable controls. Second, known as the “exclusion restriction”, exogeneity requires the instrument to affect the dependent variable only through the endogenous one which it instruments (here, the China Shock – or domestic import penetration from China through it).

In other words, the exogeneity assumption requires import from China to the US to be exogenous to unobservables determining economic outcomes (average TFP now, wages in the next sections) in Europe, and that the only channel through which the instrument affects them is via import penetration of Chinese products to France, Italy, and Spain.

The argument for the exogeneity of the instrument builds upon the idea that the rise in Chinese imports in the sample period, and especially starting from 2001, is mostly attributable to rising competitiveness, lower trade tariffs, and the 2001 China accession to the WTO. Being a supply-side shock, it should be common across all countries. Then, looking at import changes in the US (a large and perhaps heterogeneous economy) allows us to isolate the exogenous part in the domestic variation of import penetration.

b.

We re-estimate the impact of the average China Shock (1995-2006) on the average post-crisis TFP (2014-2017) using a two-stage least squares (2SLS) procedure. We instrument the China Shock index using the instrumental variable built before, based on changes in Chinese imports to the USA. The results are reported in column (2) of Table X.

The estimate on the China Shock coefficient remains significant at the 99% confidence level, and virtually unchanged in size. One standard deviation increase in the China Shock increases, on average, the mean TFP by 0.0995, or 4.46% of the sample mean. The first stage is also positive, relevant and significant, and the corresponding Kleibergen-Paap "F Statistic" is around 400, well above the rule-of-thumb value of 10 (Staiger and Stock, 1997).

Assuming that our instrument is valid, so that is also satisfies the untestable randomness assumption (i.e., uncorrelation with potential outcomes) and the exclusion restriction (it should affect the outcome only through the first stage), this result shows that the OLS estimate previously computed is not driven by domestic demand, but rather by an exogenous, supply-side shock from China. Yet these assumptions are untestable and could be violated. For instance, the randomness assumption would fail if product demand shocks were correlated between US and France, Italy, or Spain (as similarly pointed out by Autor et al., 2013). Excluding from the analysis industries with large historical correlated shock between rich countries may support the case for the exogeneity of the instrument. The exclusion restriction may fail, in turn, if exports from China to the US imply a decrease in US imports from Europe, with a direct impact on economic outcomes through trade flows with the US rather than with China.

c.

Column (3) of Table X reports the results of a linear regression model absorbing Country and industry Fixed Effects. Including relevant controls, this OLS regression of the average post crisis (2014-2017) wage on the average regional level China Shock (1995-2006) shows a coefficient of 34.22, significant at the 99% confidence level. As in point a., multiplying this coefficient for the standard deviation of the China Shock variable we get the perhaps more interpretable figure of the average impact of one extra standard deviation in the China Shock on the average TFP: 2.69, or 7.12% of the mean of the average post-crisis wages in the sample.

The estimated coefficient is thus a positive and quite sizable one, implying that to an extra standard deviation of China shock, would correspond an average annual increase in average wages of 2,690 Euro per worker. As in the case of TFP, the causal interpretation of the coefficient is not straightforward. The estimate may be biased if National import shocks, underpinning the China Shock index, are driven by (endogenous) domestic demand shocks. These domestic demand factors, perhaps unobserved, could in turn drive both wage growth and demand for Chinese imports, leading to an upwards bias in the OLS estimate, as pointed out earlier.

Then, if demand side factors drive imports from China, the OLS estimate will be biased by endogeneity, and resorting to the exogenous instrument previously discussed allows to identify and isolate the exogenous part of the China shock and estimate its impact on domestic outcomes.

Looking at the 2SLS coefficient, its value of 32.49 hints at a slightly lower impact, at least in qualitative terms, of import-competition from China on average wages, with an annual increase of 2,550 euros (6.8% increase). Again, it must be noted that the difference between the OLS and the 2SLS estimates is not statistically significant.

In economic terms, this result seems in line with the existing literature on trade shocks. The literature is almost unanimous on the negative impact of import competition from China on labour market outcomes, in particular employment, for the most exposed US industries (Autor et al. 2013, 2015; Acemoglu et al. 2016). The evidence is less clear-cut for Europe, although some evidence of negative labour market effects due to trade shock by China in Spain (Donoso et al., 2015) and Germany

(Dauth et al., 2014) exists. Evidence on the effects of import shocks from low-income countries on wages is even less conclusive. Acemoglu et al. (2016) finds a small but significant increase in wages in production workers, which is suggestive of trade-induced changes in the composition of employment. Autor et al. (2014) finds instead a decrease in wages for all workers, with high earning workers incurring modest losses and lower-wage workers being the most adversely affected. Despite further investigation is needed to shed light on the heterogeneous dynamics which might lead to an overall increase in average wages following a period of exposure to import competition from low-income countries, both the existing literature and our results suggest that there might be a skill-biased change in the composition of the labour force driven by technology at play. In this setting, low-wage and high-wage workers have, respectively, the role of the losers and the winners also when it comes to earnings (rather than employment), in line with the existing evidence on gains and losses from trade and globalisation on labour market outcomes (Autor et al. 2014, Autor et al. 2016)

d.

The significant effect of the China Shock measure on average wages, analysed in section c. and outlined by the results in column (3) and (4) of Table X, disappear in column (5) and (6) where we control for average TFP. Adding average TFP in post-crisis period leads to a loss of statistical significance of the China Shock coefficient in explaining the variation in mean wages, both in the OLS and 2SLS analysis. The magnitude of the coefficient is also strongly affected, going down from 34.22 to 7.455 in the OLS case, and from 32.49 to 5.914 for the 2SLS estimates.

This result could be due to a mediating effect of TFP in labour market outcomes following a trade shock. One of the possible consequences of the increase of import competition from China is an attempt, on behalf of the impacted firms, to increase the level of productivity to reduce production costs and to be more capable to face competition from China. This can be attained through an increase in investments in technology and automation.

Our mediation hypothesis can be tested by first looking first at the relationship between TFP and China Shock in column (1) and (2), at the relationship shown in columns (3) and (4) between China Shock and mean average wages, and at the correlation between mean average wages and TFP, for which we report the correlation in Table XI, shows a correlation of 0.561.

Table XI

Matrix of correlations

Variables	(1)	(2)
(1) Mean TFP	1	
(2) Mean Wages	0.561*	1

The correlational evidence above points to a case of Omitted Variable Bias in the coefficients of China shock in columns (3) and (4) of Table X, where the omitted variable is average TFP. Moreover, together with the loss of significance of the China Shock coefficient, it suggests a *complete mediation* of TFP for mean wages.

Problem VII

a. (on do file)

Table XII
INDIVIDUAL-LEVEL ESTIMATES:
IMPACT OF THE AVERAGE CHINA SHOCK (1995-2006) ON THE
PROBABILITY OF VOTING FOR RADICAL-RIGHT PARTIES IN ITALY

	(1)	(2)	(3)	(4)
Estimation Method	OLS	OLS	2SLS	2SLS
Dependent Variable		**Radical Right Dummy**		
China shock	0.260***	0.174**	0.261***	0.204***
	(0.0671)	(0.0761)	(0.0588)	(0.0767)
Age	0.000778***	0.000560***	0.000778***	0.000550***
	(0.000223)	(0.000190)	(0.000220)	(0.000183)
Female	-0.0206***	-0.0214**	-0.0206***	-0.0214***
	(0.00636)	(0.00756)	(0.00620)	(0.00735)
Observations	2,596	2,005	2,596	2,005
R-squared	0.021	0.015	0.021	0.015
Education Dummies	Yes	Yes	Yes	Yes
Excluding Lom. and Ven.	No	Yes	No	Yes
First Stage Results				
IV China shock	-	-	0.0731***	0.0727***
	-	-	(0.00211)	(0.00369)
F-statistic instruments			1196	388.9

Individual post-stratification weights including design weight
Robust standard errors in parentheses, clustered at the Nuts2 level
*** p<0.01, ** p<0.05, * p<0.1

b.

Column (1) of Table XII shows the impact of exposure to the China Shock on the probability of having voted for a Radical Right party (Lega Nord or Fratelli d'Italia) in the 2013 Italian National Elections. The simple OLS estimate for the China Shock, controlling for gender, age, dummies for the (ISCED) level of education, and clustering standard errors at the regional level, is positive and significant. Considering the weighted average (using post-stratification sample weights) of the Radical Right Dummy of 0.0397, this effect is also quite large: on average, for every standard deviation increase of the China Shock the probability of having voted for the radical right rises by 2 percentage points, or 51% of the baseline probability.

Such large effect is significant under standard errors clustered at the regional level. Yet, clustering is not free of concerns and may induce small-sample bias. While clustering at the regional level is necessary to control for within-region correlations, which would lead to misleadingly small standard errors, large t-statistics, and flawed p-values, the `cluster` option in Stata computes a generalization of the White-Eicker robust covariance matrix that is consistent only for a large number of groups. Moreover, clustering reduces the effective sample over which standard errors are computed, inducing small-sample biases despite the large dataset at hand. Here, we only observe 20 groups (regions) across which values of the China Shock variable may vary, and within some of them the number of individuals can be worryingly low for statistical inference to be accurate[1].

Moreover, and especially so with Italy as the only country under consideration, our estimate could be driven by spurious correlations between historical voting patterns and regional employment in the manufacturing industry, especially in sectors more affected by the China Shock. This could be the case for Lega Nord, Italy's north quasi-secessionist party at the time, and thus historically strong in Lombardy and Veneto – two of the regions most affected by the China Shock – for reasons unrelated to the shock itself, as testified by the voting history of the two regions. This anticipates China's entry in the WTO as well as most of the years included in the computation of the China Shock index (Eligendo – Ministero dell'Interno). The problem is exacerbated by the fact that, restricting our analysis to Italy, we only observe 20 values of the China Shock variable, and these two possibly spurious correlations may be driving most of the estimate. Therefore, we assess the robustness of the model to this possible source of bias by re-estimating the same regression excluding Lombardy and Veneto column (2) Table XII. We note how this exclusion exacerbates the small-sample bias discussed at the beginning of this section. Remarkably, although the coefficient drops by 33%, results are still large and significant, hence robust to the exclusion of Lombardy and Veneto. This increases our confidence in the estimate. Concerns related to historical persistence and few observations still apply, and could be accounted for by enlarging our sample and/or by using a repeated cross-section to introduce region or country fixed effects to control for historical voting patterns which differ across regions and for common political trends within a country.

Other concerns arise from the possible endogeneity of the China Shock index itself. Possibly, regions closer to the mainstream (incumbent) parties may lobby the central government to obtain industry-specific protectionist trade policies which shield their economic activity from Chinese competition. Voters in such regions would then face a smaller import shock and be more inclined to vote for the incumbents, but for reasons related to the omitted variable of historical political ties, rather than to changes in imports. As a result, the OLS could be upwardly biased. Yet, with tariffs and non-tariff barriers set at the European level, it is unlikely that a single region could successfully lobby the EU into targeted protectionist measures that could be detrimental to the single market.

Another possible source of bias are *income effects* inducing positive demand shocks. In that case, the increase in imports from China would stem from an overall increase in demand in the Italian economy, spilling over to foreign (and Chinese) products. In this case, an increase in the import of Chinese goods would be correlated with an increase in support for the incumbent. Contrary to the previous case, here the OLS estimate would be biased downwards.

c.

To correct for the aforementioned endogeneity issues, we resort again to an IV approach, exploiting US import shocks from China as in instrument to isolate exogenous part in the Italian ones. Chinese imports to the US should be uncorrelated with demand shocks in Italy, and hence capture the exogenous variation in Chinese imports to Italy. More precisely, demand shocks in the US should be uncorrelated with potential electoral outcomes in Italy (randomness assumption) and affect them only

[1] One way to correct for this, would be to compute standard errors through cluster bootstrapping with asymptotic refinement, which compute standard errors by randomizing the bootstrapping procedure at the cluster level. We do not engage in such sorcery.

through the channel of their correlation with the exogenous variation in Chinese imports to Italy (exclusion restriction).

The rationale underpinning the exogeneity of the outcome is perhaps more convincing in this case than when looking at economic outcomes, as the exogeneity of trade imports in the US to political outcomes in Italy is perhaps more intuitive than that of economic outcomes, where common productivity or technology shocks may be at play, but we can still think of circumstances where the exclusion restriction would fail. For instance, a rise in imports from China to the US may affect political preferences there, as shown by Autor et. al (2020), and in turn influence political trends in Italy. To the extent that communication and cultural influxes are frequent, politics in Italy and the US may be partially intertwined, and political trends may spill over from the US to Italy. If this were the case, and especially if political trends in the US preceded those in Italy, we would have correlated political shocks across countries which undermine the exclusion restriction as the instrument affects political outcomes in Italy through a channel other than trade exposure. Existing empirical evidence supports the existence of political contamination across countries (Dominioni et al. 2020). Moreover, this channel is perhaps more prominent in the long run, as in the case for the outcome under consideration here. Yet, to the extent that this effect is marginal and that imports in the US predict well those in Europe (relevance of the instrument), the identification scheme remains overall strong.

Instrumenting the China Shock by the instrument built previously, our results remain virtually unchanged. The first stage in column (3) of Table XII is positive, significant and nonnegligible in size, with a Kleibergen-Paap “F Statistic” of 1196, again well above the rule of thumb value of 10 (Staiger and Stock, 1997). On one hand this could be evidence of the strength of the instrument, on the other, very high values of the F-Statistic may raise concerns related to collinearity of the endogenous variable and the instrument in our sample and rise small sample bias concerns. Again, the instrumented model is robust to the exclusion of regions with a historically large vote share for Lega Nord (Lombardy and Veneto), although the estimate on the China Shock variable drops by 21%. The F-stat falls as well although remaining well above the thumb rule of 10.

d.

Comparing the OLS estimate in column (1) of Table XII with the 2SLS one in column (3) we point out how the two are strikingly similar. The main reason for this is twofold.

On one hand, the lack of bias in the OLS estimate with respect to the IV one points to a fully exogenous shock in the Italian imports from China, driven by supply-side factors in China that could then be precisely estimated looking at import shocks in the US.

On the other hand, as with the original China Shock index, the instrumented one is also particularly exposed to small sample biases under our clustering restrictions. As underlined in b., it is then likely that, by looking at sheer voting shares in a cross-section, and with just 20 groups over which the China Shock index can vary, we fail to account for historical voting patterns in few regions which drive the overall effect while possibly being in part spuriously correlated with the instrumented China Shock just as they were with the original index.

REFERENCES

Altomonte, Carlo, Coali, Andrea. 2020. Employment, Productivity and Import Shock: Evidence from the European Manufacturing Industry. *MICROPROD Raising EU Productivity: Lessons from Improved Micro Data Supply and Demand-Oriented Economic Policies to Boost Robust Growth in Europe.* http://www.microprod.eu/wpcontent/uploads/2020/03/D5_2_Altomonte_Coali.pdf

Ministero dello Sviluppo Economico. 2017. *Relazione Annuale Al Parlamento Sullo Stato D'Attuazione E L'Impatto Delle Policy a Sostegno Di Startup E PMI Innovative.* https://www.mise.gov.it/images/stories/documenti/startup_relazione_annuale_al_2017.pdf

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. *Identification Properties of Recent Production Function Estimators.* Vol. 83. doi:10.3982/ECTA.

Auer, Raphael A., Kathrin Degen, and Andreas M. Fischer. 2013. "Low-Wage Import Competition, Inflationary Pressure, and Industry Dynamics in Europe." *European Economic Review* 59: 141-166. doi:10.1016/j.euroecorev.2012.11.007. https://dx.doi.org/10.1016/j.euroecorev.2012.11.007.

Autor, David H., David Dorn, Gordon H. Hanson, and Jae Song. 2014. "Trade Adjustment." *The Quarterly Journal of Economics* 129 (4): 1799-1860. doi:10.1093/qje/qju026. https://www.jstor.org/stable/26372587.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *The American Economic Review* 103 (6): 2121-2168. doi:10.1257/aer.103.6.2121. https://www.jstor.org/stable/42920646.

Autor, David, David Dorn, and Gordon H. Hanson. 2016. *The China Shock: Learning from Labor Market Adjustment to Large Changes in Trade.* Cambridge, Mass: National Bureau of Economic Research. doi:10.3386/w21906. http://www.nber.org/papers/w21906.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi. 2020. "Importing Political Polarization? the Electoral Consequences of Rising Trade Exposure." *The American Economic Review* 110 (10): 3139-3183. doi:10.1257/aer.20170011. https://search.proquest.com/docview/2444851190.

Beveren, Van. a. *A Service of Zbw Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics Total Factor Productivity Estimation: A Practical Review LICOS.*

Van Beveren Ilke. 2007 *Total Factor Productivity Estimation: Total factor productivity estimation: A practical review, LICOS Discussion Paper, No. 182, Katholieke Universiteit Leuven, LICOS Centre for Institutions and Economic Performance, Leuven*

Bloom Nicholas, Draca Mirko, Van Reenen John. 2016. *Trade Induced Technical Change? the Impact of Chinese Imports on Innovation, IT and Productivity* Oxford University Press.

Cobbold Trevor. 2003. *A Comparison of Gross Output and Value-Added Methods of Productivity Estimation.* Australian Government, Internal Working Document.

https://www.pc.gov.au/research/supporting/comparison-gross-output-value-added-methods/cgovam.pdf

Colantone, Italo And Piero Stanig. 2018. "Global Competition and Brexit." *The American Political Science Review* 112 (2): 201-218. doi:10.1017/S0003055417000685. https://dx.doi.org/10.1017/S0003055417000685.

Colantone, Italo and Piero Stanig. 2018. "The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe." *American Journal of Political Science* 62 (4): 936-953. doi:10.1111/ajps.12358. https://www.jstor.org/stable/26598793.

Dominioni, Goran, Alberto Quintavalla, and Alessano Romano. 2020. "Trust Spillovers among National and European Institutions." *European Union Politics* 21 (2): 276-293. doi:10.1177/1465116519897835. https://www.narcis.nl/publication/RecordID/oai:repub.eur.nl:124153.

Donoso, Vicente, Víctor Martín, and Asier Minondo. 2015. "Does Competition from China Raise the Probability of Becoming Unemployed? an Analysis using Spanish Workers' Micro-Data." *Social Indicators Research* 120 (2): 373-394. doi:10.1007/s11205-014-0597-7. https://doi.org/10.1007/s11205-014-0597-7.

Gandhi, Amit, Salvador Navarro, and David A. Rivers. 2020. "On the Identification of Gross Output Production Functions." *The Journal of Political Economy* 128 (8): 2973-3016.doi: 10.1086/707736. https://www.journals.uchicago.edu/doi/abs/10.1086/707736.

Hoch, Irving. 1962. "Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data." *Econometrica* 30 (1): 34-53. doi:10.2307/1911286. http://www.jstor.org/stable/1911286.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous Univariate Distributions, Volume 1* Wiley. https://books.google.it/books?id=AqBOswEACAAJ.

Marschak, Jacob and William H. Andrews. 1945. "Random Simultaneous Equations and the Theory of Production." *Econometrica* 13 (1): 91. doi:10.2307/1907137. https://www.jstor.org/stable/1907137.

Massey, Frank J. 1951. "The Kolmogorov-Smirnov Test for Goodness of Fit." *Journal of the American Statistical Association* 46 (253): 68-78. doi:10.2307/2280095. http://www.jstor.org/stable/2280095.

Melitz, Marc. 2003. *The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity.* Cambridge, Mass: National Bureau of Economic Research.

Melitz, Marc and Gianmarco Ottaviano. 2008. *Market Size, Trade, and Productivity*: Review of Economic Studies. doi:10.3386/w11393. http://www.nber.org/papers/w11393.

Ministero dell'Interno. "Eligendo - Archivio Elezioni.", https://elezioni.interno.gov.it/.

Mundlak, Yair. 1961. "Empirical Production Function Free of Management Bias." *American Journal of Agricultural Economics* 43 (1): 44-56. doi:https://doi.org/10.2307/1235460. https://doi.org/10.2307/1235460.

Olley, G. S. And A. Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 1263-1297. doi:10.2307/2171831. https://www.jstor.org/stable/2171831.

Petrin, Amil and James A. Levinsohn. 2003. *Estimating Production Functions using Inputs to Control for Unobservables*: The Review of Economic Studies. doi:10.3386/w7819. http://www.nber.org/papers/w7819.

Stock, James H. and Douglas O. Staiger. 1994. *Instrumental Variables Regression with Weak Instruments*. Cambridge, Mass: National Bureau of Economic Research. doi:10.3386/t0151. http://www.nber.org/papers/t0151.